Probability and Statistics

Best Practices in Dam and Levee Safety Risk Analysis

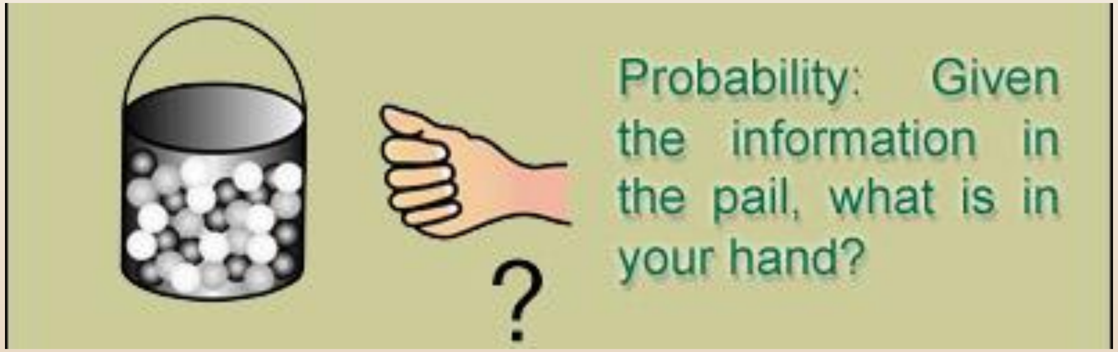
Part A – Risk Analysis Basics

Chapter A-1

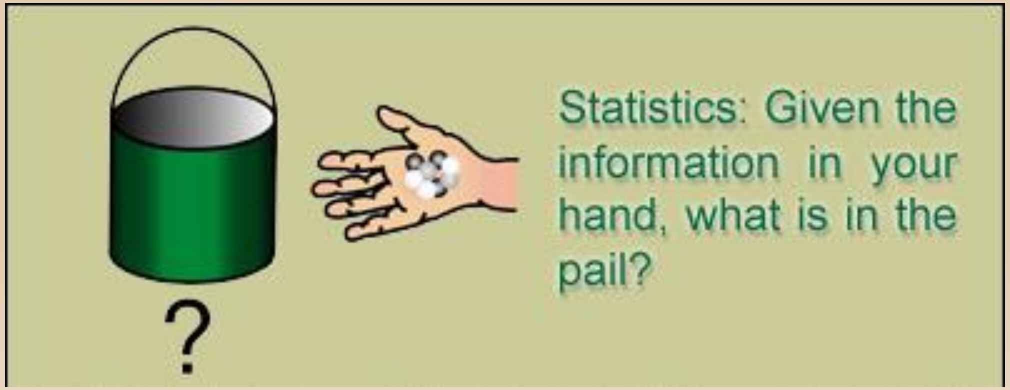
July 2018

US Army Corps of Engineers®

Objectives



- Define terms
- Develop theory
- Demonstrate common applications



Outline of Topics

- Sets Theory and Events
- Probability Theory
- Statistics
- Monte Carlo Method

"I can prove it or disprove it! What do you want me to do?"

Key Concepts

- **Risk Analysis** utilizes *Set Theory*, *Probability Theory*, and *Statistics* to improve and communicate our understanding of the risks associated with the operation of water retention infrastructure.
- **Set Theory** provides a framework for the analysis of events and the relationships between events. It is based on logic alone.
- **Probability Theory** provides a framework for analyzing the *likelihoods* of events and combinations of events. It is based on set theory and math.
- **Statistics** is the branch of science that deals with the interpretation and analysis of data. The concepts of distributions and random variables are introduced through statistics.

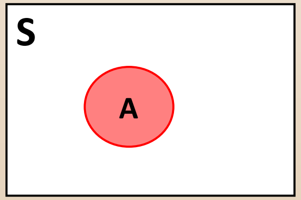
Set Theory

- Used to describe relationships between events
- **Events** are the basic building blocks of risk analysis. One way to describe an **event** is as something that could happen, projected into the present as certain (no matter how likely or unlikely)
- e.g. “An earthquake could occur in September” can be projected into the present as “an earthquake occurs in September”
- Other examples of events:
  - Joe attends Best Practices training
  - Bill misses his flight to Denver and misses his 5 AM alarm
  - The maximum reservoir elevation in 2018 exceeds El. 4453

Set Theory

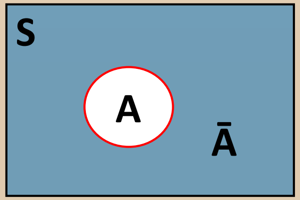
Sample Space

Set of possible events



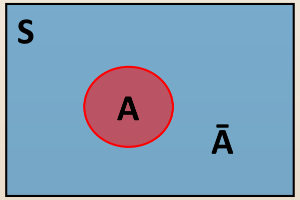
Event (A)

The levee overtops



Complement (Ā)

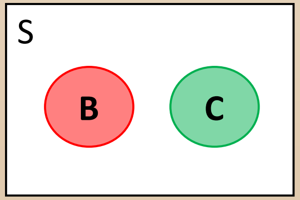
"Not"
The levee does not overtop



Collectively Exhaustive

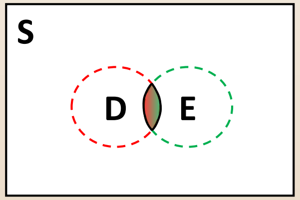
Covers the entire sample space.

The levee overtops or it does not overtop.



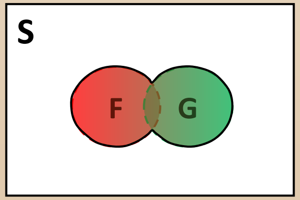
Mutually Exclusive

Both cannot occur.
One gate opens.
Two gates open.



Intersection (D ∩ E)

"And".
A person does not evacuate and the flood depth at their location is greater than 15 feet

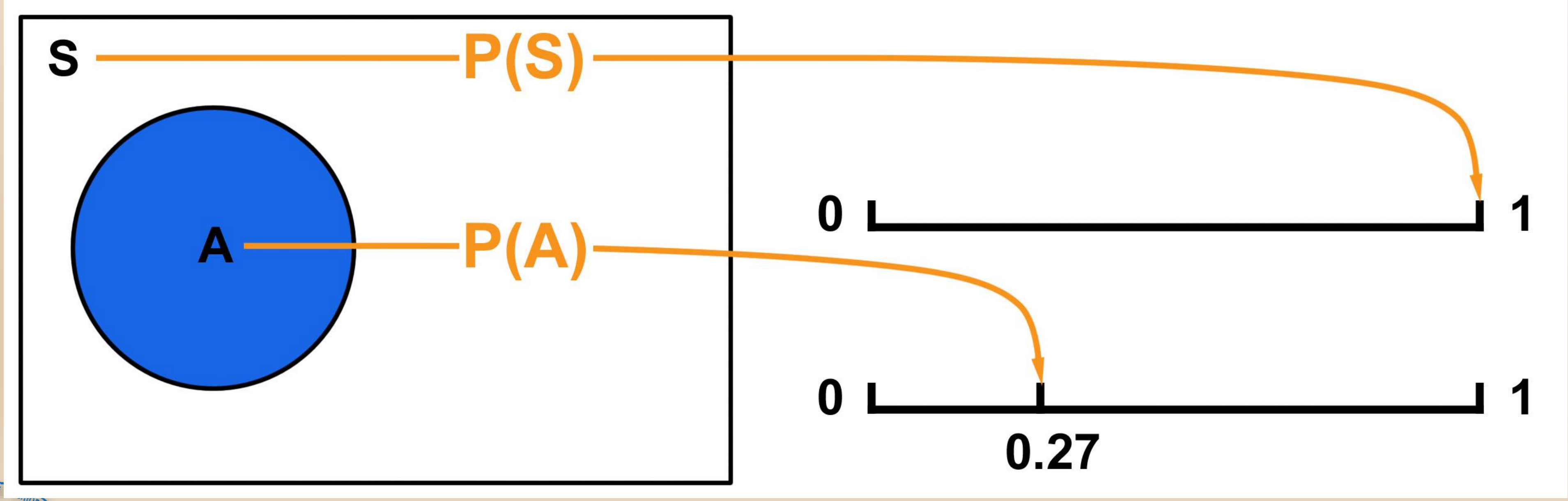


Union (F U G)

"Or"
A spillway monolith slides or internal erosion initiates in the embankment

Probability Theory

- Probability theory introduces the concept of “size” to the sample space.
- p(...) can be thought of as a function that maps events in sample space to the real number interval [0,1]



Probability Axioms

- Probabilities are non-negative real numbers

$$P(A) \geq 0$$

- Probability of the certain event is 1.0

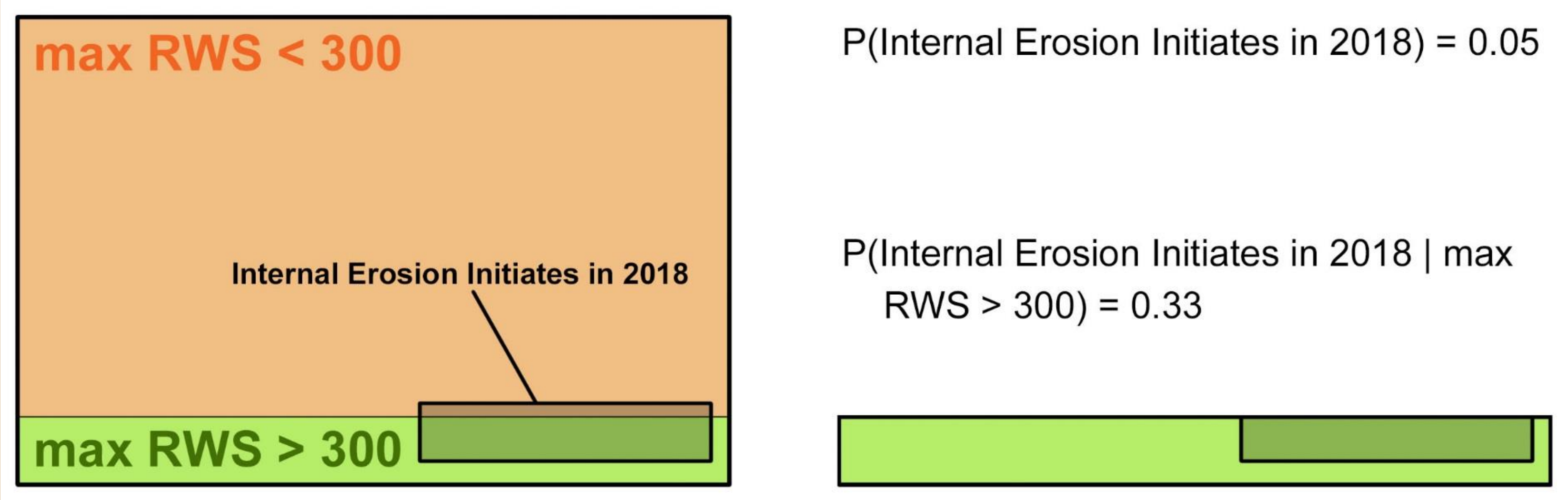
$$P(S) = 1$$

- Probability of the union of two mutually exclusive events is equal to the sum of their probabilities

$$P(A \; U \; B) = P(A) + P(B)$$

Conditional Probability and Statistical Independence

- The expression P(A|B) is read “probability of A given B”
- Example: Probability that internal erosion initiates vs. probability that internal erosion initiates <u>given</u> that the maximum annual reservoir elevation exceeds El. 300



- If P(A|B) = P(A), then events A and B are statistically independent

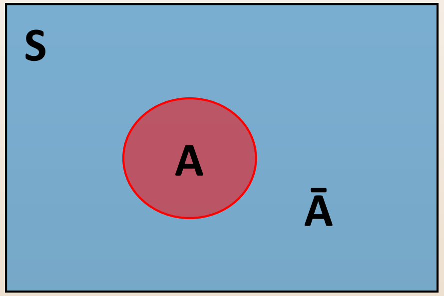
Probability Interpretations

- Physical (frequency)
  - Basis: observation
  - E.g., spillway flow has occurred twice over the past decade. The annual probability of spillway flow is about 0.2.

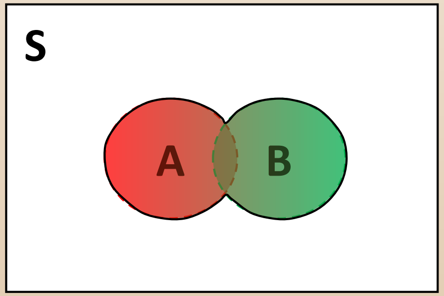
- Evidential (degree of belief)
  - Basis: Personal experience, expert judgment, weight of evidence
  - E.g., probability of cracking is about 0.3% based on construction records, measured concrete strengths, the results of an NL finite element analysis, and observed performance of similar dams

Commonly used probability formulas

- $P(\bar{A}) = 1 - P(A)$
  - Probability of no breach ($\bar{A}$) equals one minus the probability of breach (A)



- $P(A \cup B) = P(A) + P(B) - P(AB)$
  - Total probability of breach given two potential failure modes, A and B

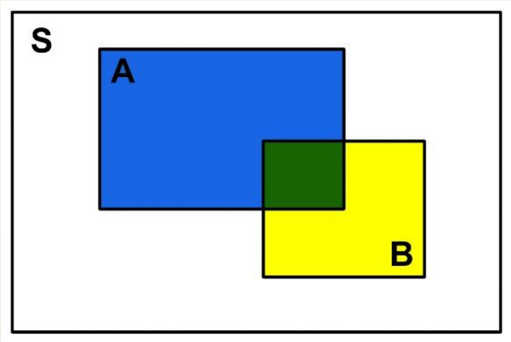


- Total probability theorem
  - For a set of mutually exclusive and collective exhaustive events E_i
  - $P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \ldots + P(A|E_n)P(E_n)$

Commonly used probability formulas

$P(A \cap B) = P(A) * P(B \mid A)$

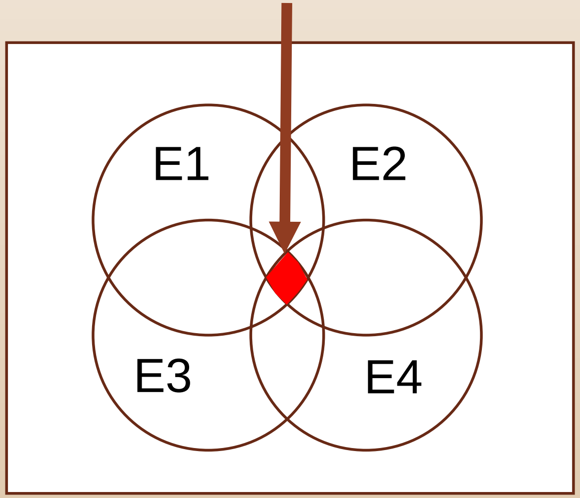
- General case – applies to all types of intersections
- Basic formula used to in risk analysis
- e.g. the probability of a breach occurring in an earthquake is equal to the probability of the earthquake times the conditional probability of failure given the earthquake



- $P(A \cap B) = P(A) * P(B)$
  - Special case - statistically independent events only (the events that define a given potential failure mode are typically not statistically independent)

Example:
Probability theory applied to risk analysis

- What is the annualized probability of failure?
- Assuming failure results from a sequence of four events **E1E2E3E4**
- An earthquake occurs, P(E1)
- Foundation liquefaction occurs, P(E2|E1)
- The crest settles, P(E3|E1 ∩ E2)
- The dam overtops, P(E4|E1 ∩ E2 ∩ E3)
- The annualized failure probability is calculated as the probability of the intersection
- AFP = P(E1) * P(E2|E1) * P(E3|E1 ∩ E2) * P(E4|E1 ∩ E2 ∩ E3)



Binomial Theorem
Pascal's Triangle

- Probability of *k* occurrences in *n* trials

$$\binom{n}{k} p^k (1-p)^{n-k}$$

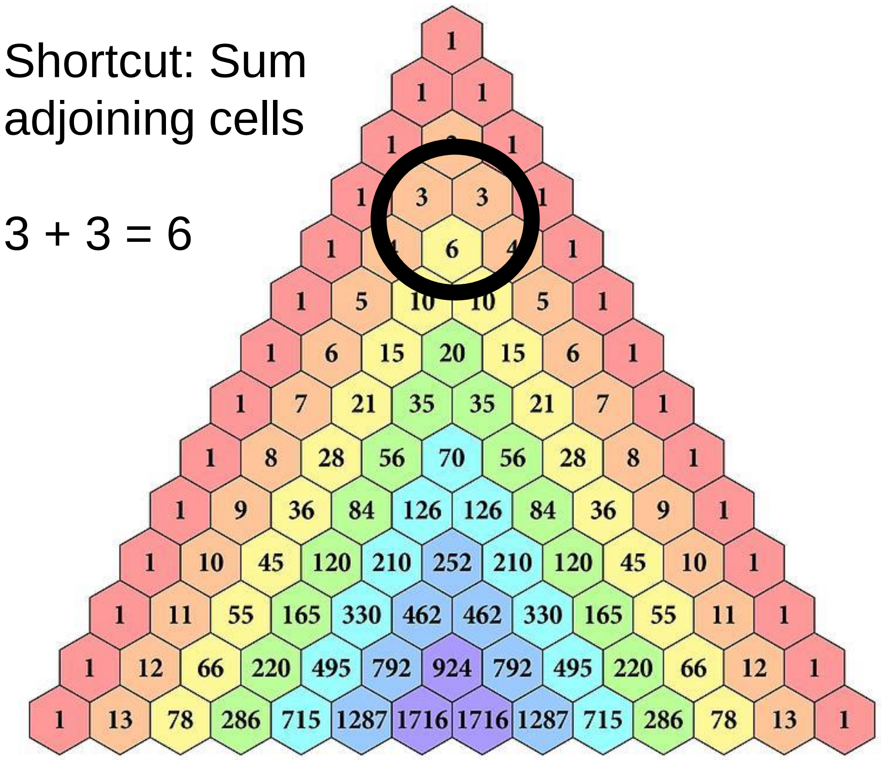
- Trials must be statistically independent
- Given 4 spillway gates each with 0.1 probability of not opening
- Probability of 2 gates not opening is

$$(6)(0.1)^1(1-0.1)^3 = 0.05$$

Column is k with values from 0 to n

Shortcut: Sum adjoining cells

3 + 3 = 6



Row is n with values from 0 to ∞

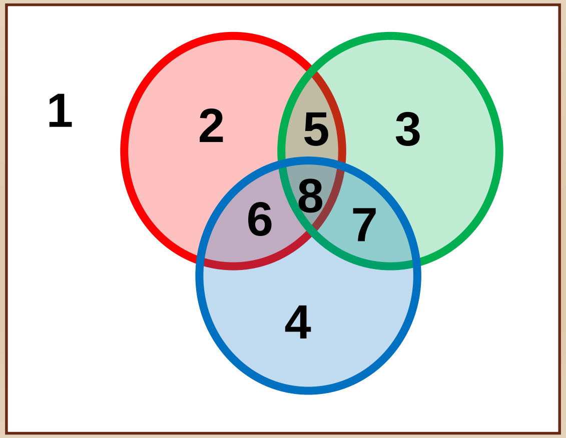
$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Combinations for Statistically Independent Events

- System of two dams and one levee
- Two outcomes for each each
  - Breach or No Breach
- $2^3 = 8$ possible combinations

- Binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$



Combination	Dam A	Dam B	Levee
1	NB	NB	NB
2	**B**	NB	NB
3	NB	**B**	NB
4	NB	NB	**B**
5	**B**	**B**	NB
6	**B**	NB	**B**
7	NB	**B**	**B**
8	**B**	**B**	**B**

1 combination for k=3, n=0

3 combinations for k=3, n=1

3 combinations for k=3, n=2

1 combination for k=3, n=3

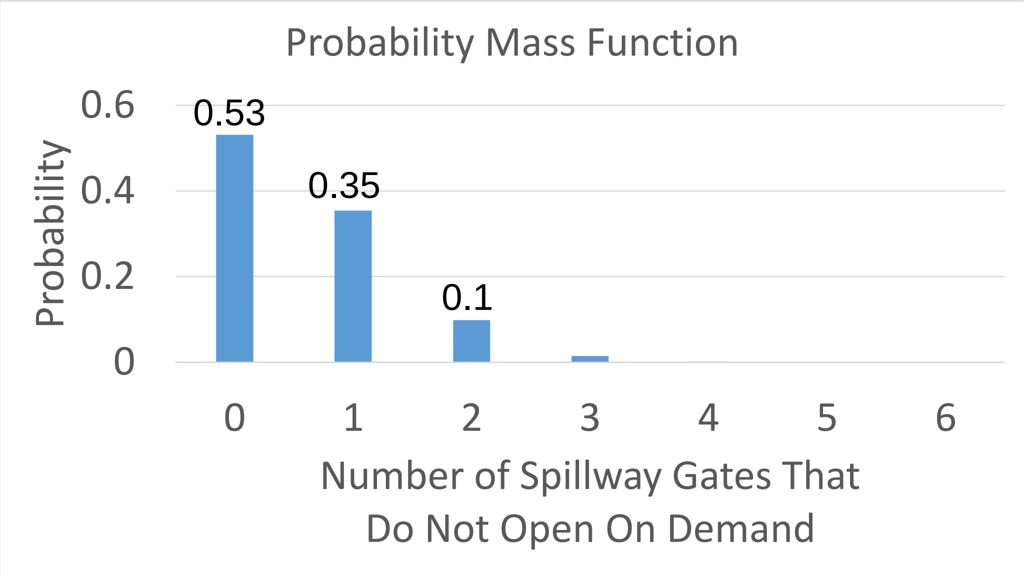
B = Breach ; NB = No Breach

Random Variable

- A random variable is used to represent a parameter whose value can take on a variety of different outcomes
- Examples in dam safety risk analysis:
  - The return period of a 1 Million cfs flood ranges from 5000 to 10,000 years
  - The exceedance probability of a 0.8g PHA ranges from 0.001 to 0.01
  - The conditional probability of internal erosion initiation ranges from 1E-4 to 1E-3
- Two basic types of random variables:
  - *Discrete RVs* have specific sets of outcomes
  - *Continuous RVs* have an infinite number of outcomes

Probability Mass / Density Functions

Discrete (PMF)

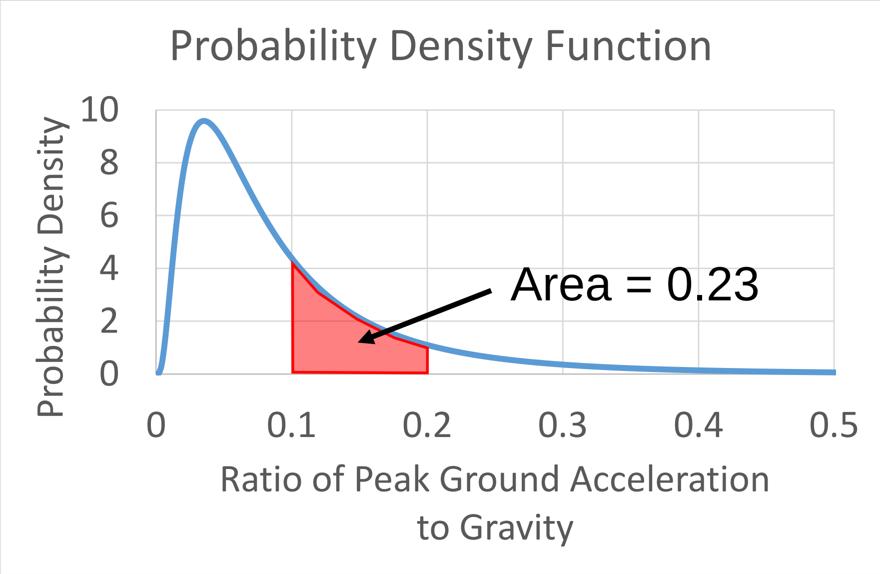


$$P(n = 2) = 0.1$$

$$P(n \leq 1) = 0.53 + 0.35 = 0.88$$

$$\sum_{i=1}^{6} P(n = i) = 1$$

Continuous (PDF)



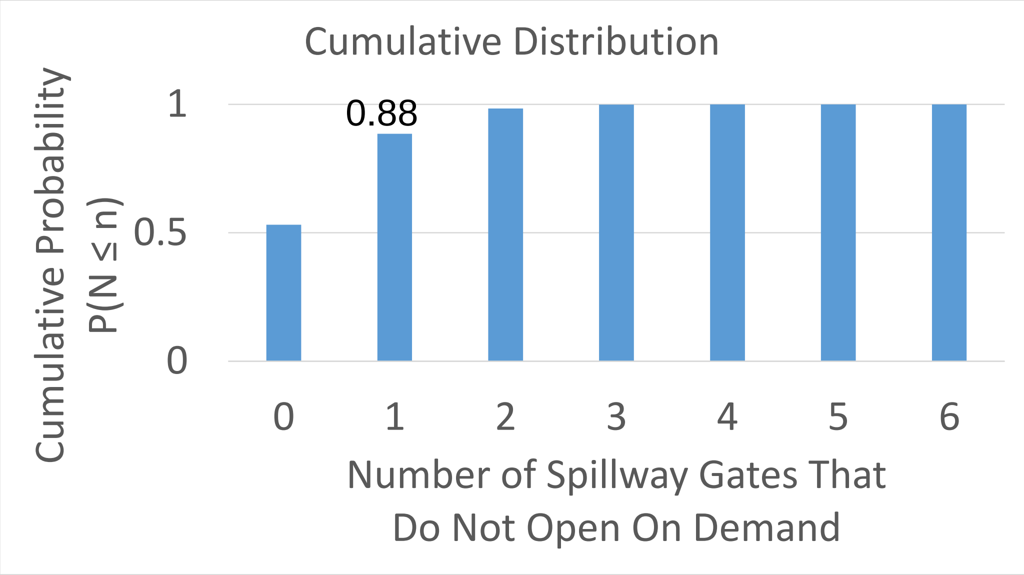
$$P(PGA = 0.1) is\ undefined$$

$$P(0.1 < PGA \leq 0.2) = \int_{0.1}^{0.2} f_x(x)dx = 0.23$$

$$\int_{-\infty}^{\infty} f_x(x)dx = 1$$

Cumulative Distribution Function

Discrete (CDF)

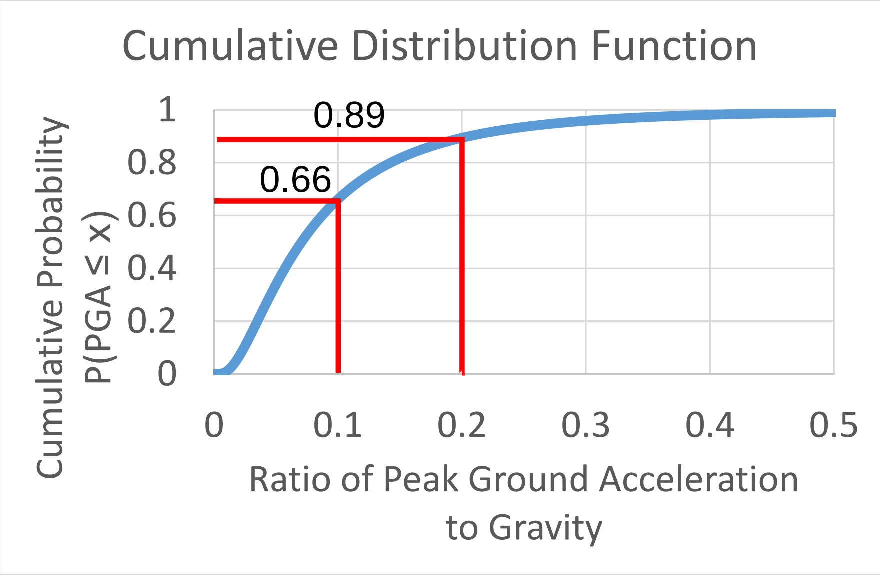


$$P(n = 2) = P(n \leq 2) - P(n \leq 1)$$
$$= 0.98 - 0.88 = 0.1$$

$$P(n \leq 1) = 0.88$$

$$P(n \leq 6) = 1$$

Continuous (CDF)



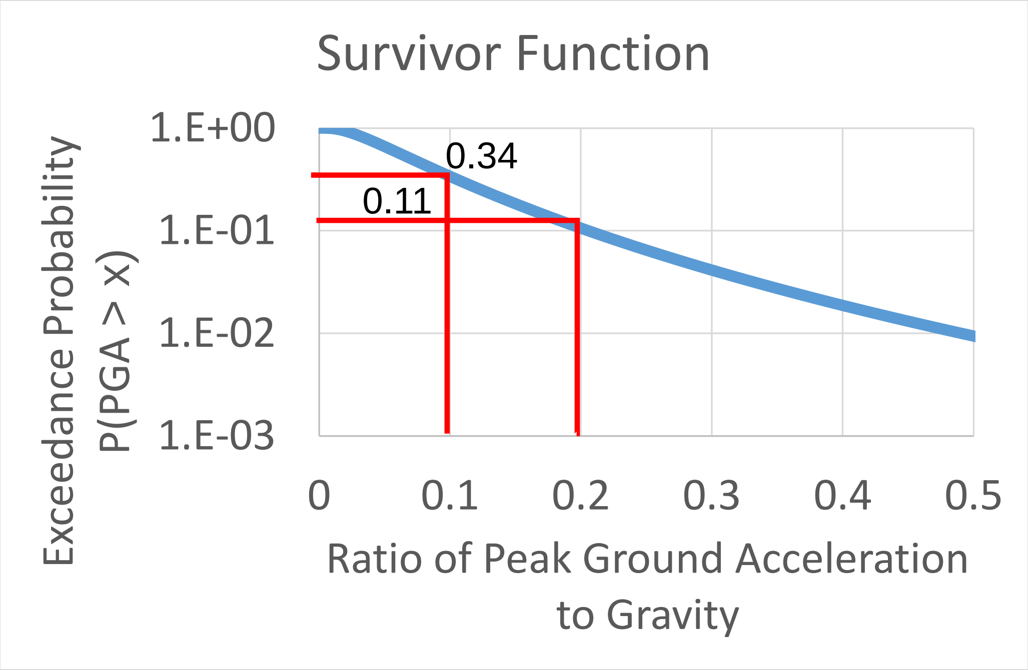
$$P(PGA = 0.1) is\ still\ undefined$$

$$P(0.1 < PGA \leq 0.2) = 0.89 - 0.66 = 0.23$$

$$P(PGA \leq \infty) = 1$$

Exceedance Curve ("survivor function")

- Complement of the cumulative distribution function (1 – CDF)
- CDF would be the non-exceedance curve
- Commonly used to portray flood and seismic hazards

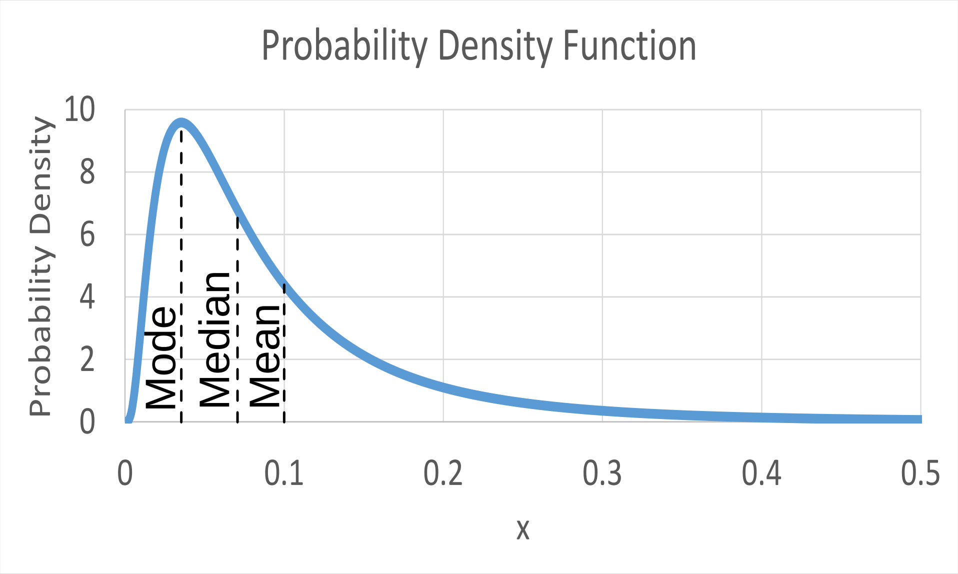


$$P(PGA > 0.2) = 0.11$$

$$P(0.1 < PGA \leq 0.2) = 0.34 - 0.11 = 0.23$$

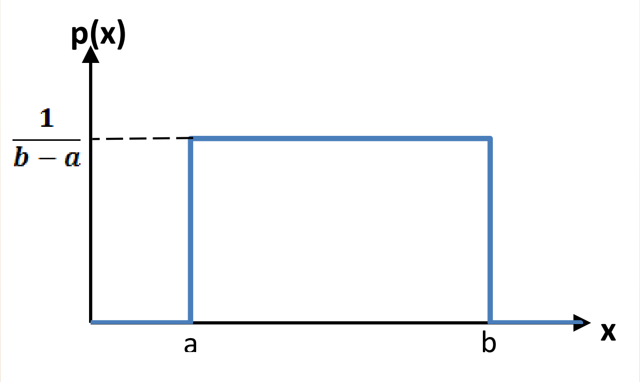
Point Estimators

- Mean – Average value
- Median – 50th percentile
- Mode – Most frequent

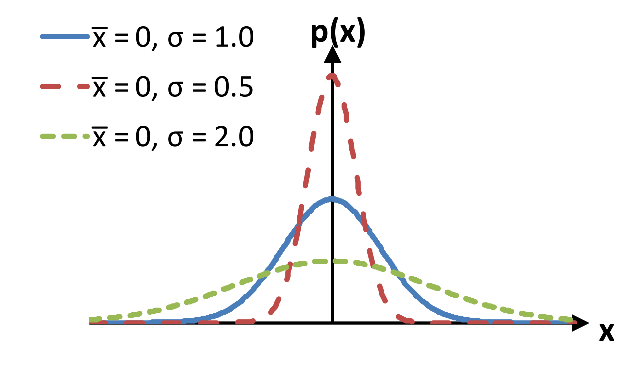


- Mean – First moment, centroid
- Variance – Second moment, spread
  - Standard Deviation $\sqrt{Variance}$
  - Coefficient of Variation $\frac{Standard\ Deviation}{Mean}$
- Skew – Third moment, symmetry

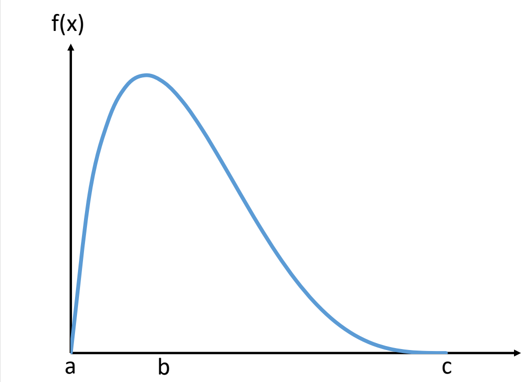
Common Probability Distributions



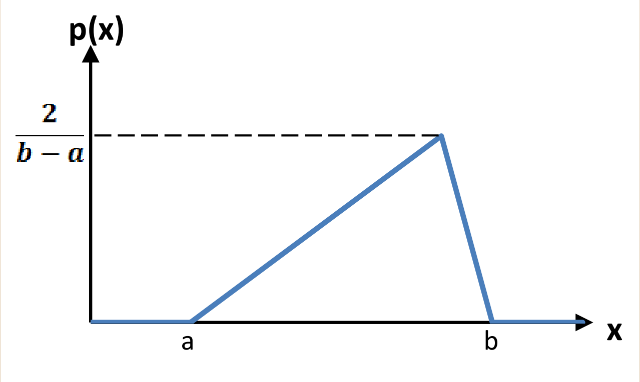
Uniform



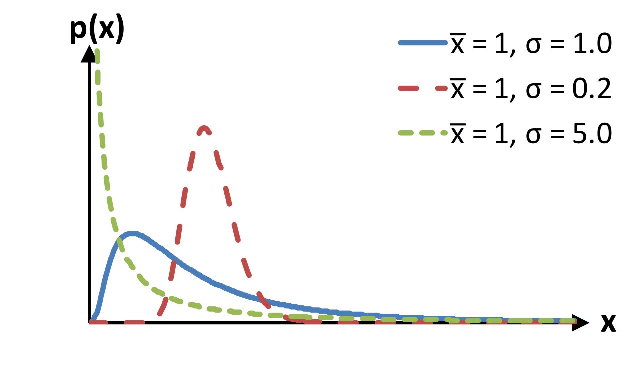
Normal



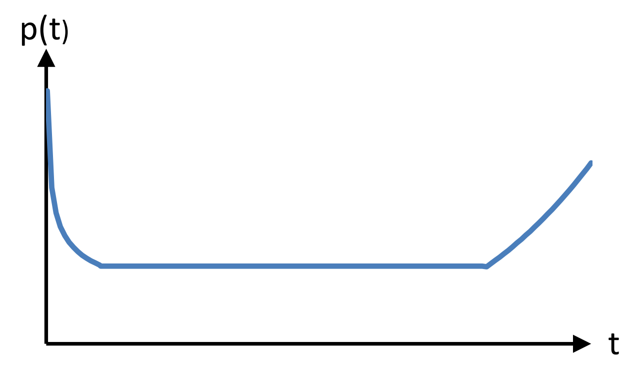
PERT



Triangular



Log-Normal



Weibull

Bayesian inference

- Observational Method
- Weight of evidence
- Value of information
- Minimize cognitive biases
- *Note: In practice, it is more common to qualitatively adjust the probability estimates for new information.*

Bayes Theorem

- P(x) : Prior probability of an event
- P(O|x) : Probability of an observation given the event
- P(O) : Total probability of the observation (normalizing constant)
- P(x|O) : Posterior probability of the event given the observation

$$P(x|O) = \frac{P(x)P(O|x)}{P(O)}$$

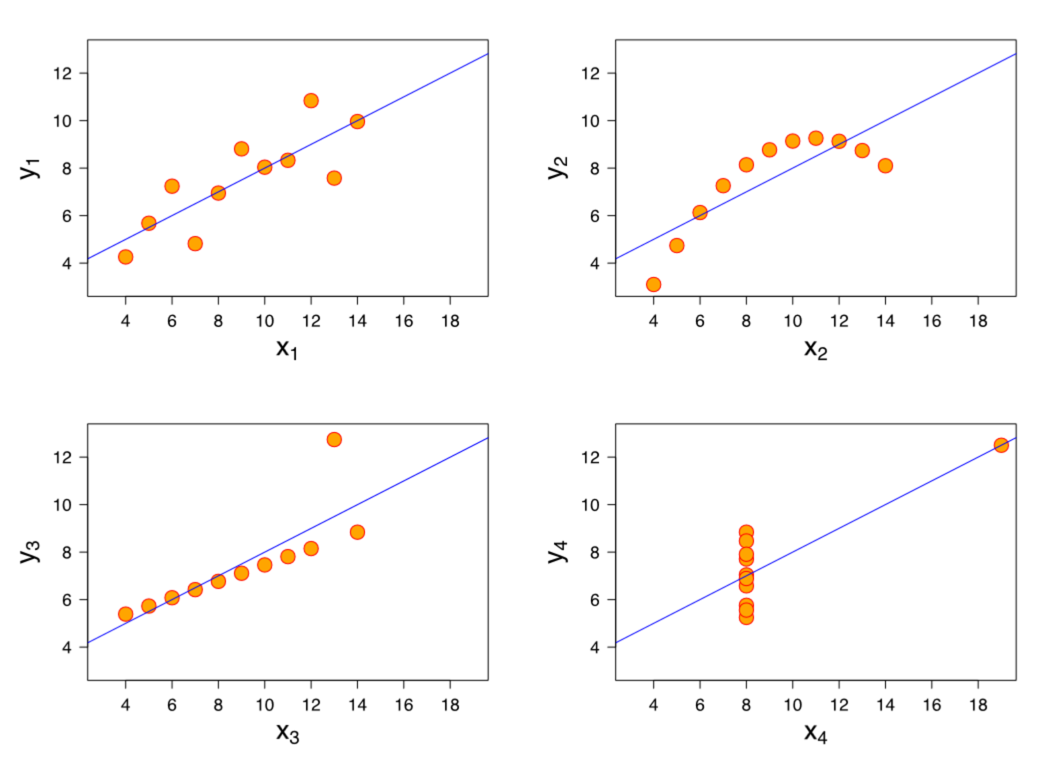
Bayes Theorem

$$P(x|O) = \frac{P(x)P(O|x)}{P(O)}$$

- P(x) : Does the levee foundation have a permeable layer
  - Initial estimate of **0.2** based on experience and judgment
- P(O|x) : Exploration does not find a permeable layer
  - Boring spacing about 500 feet; Layer extent should be about 200 feet
  - Probability of NOT finding (assuming it is there) is about 300/500 = **0.6**
- P(O) : Total probability of the observation
  - Exists AND Not Found + Does Not Exist AND Not Found
  - 0.2 * 0.6 + 0.8 * 1 = **0.92**
- P(x|O) : Does the levee foundation have a permeable layer given no layer was found
  - 0.2 * 0.6 / 0.92 = **0.13**

Correlation

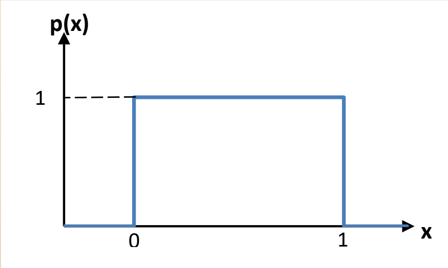
- Any statistical relationship between two random variables
  - Linear commonly used
  - Nonlinear options available
- Used for predictive relationships
  - Friction angle from SPT blow count
  - Probability of rain today from yesterday's weather



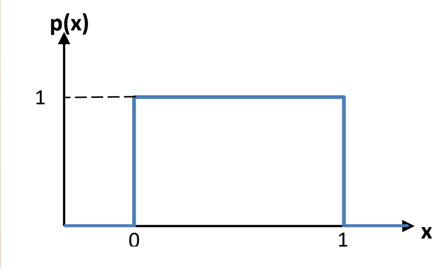
Four sets of y values with the same mean, variance, correlation coefficient, and regression line.

Central Limit Theorem

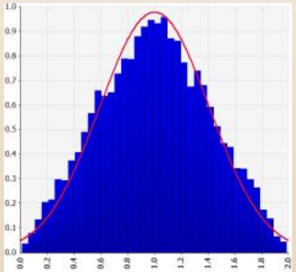
- Sum of distributions trends toward a normal distribution



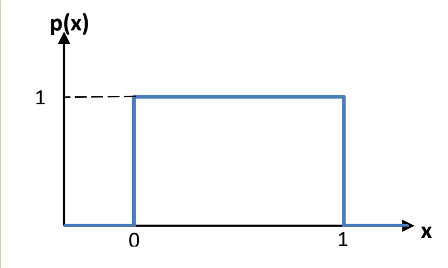
+



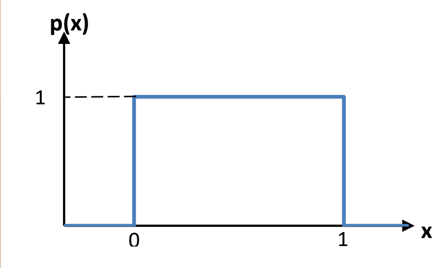
=



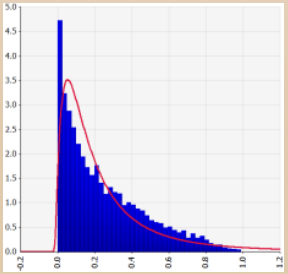
- Product of distributions trends toward a log normal distribution



x

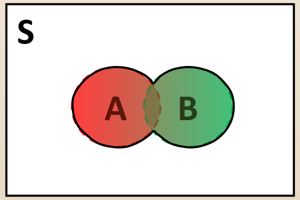


=

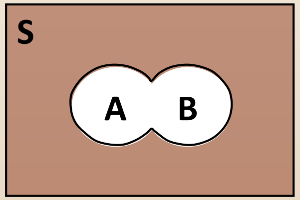


de Morgan’s Rule

- In theory
  - The complement of the union of two or more events

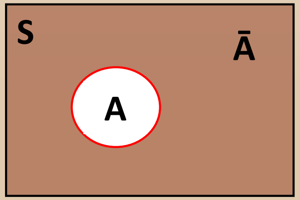


(A U B)

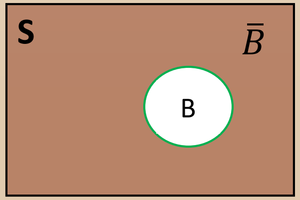


$(\overline{A \cup B})$

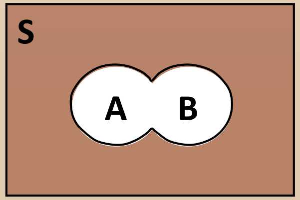
  - is equal to the intersection of their complements



Complement (A)



Complement (B)



$\bar{A} \cap \bar{B}$

- For <u>statistically independent events</u>, the union probability is sometimes easier to compute in this way

E.g., total (union) risk of failure for a facility

product of the complementary probabilities of the individual PFMs

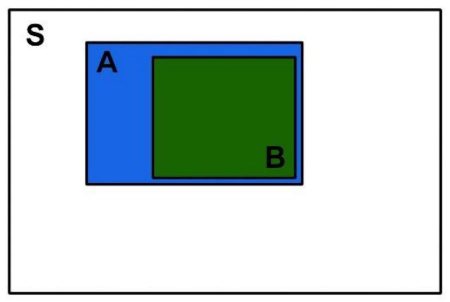
$$P\,(f) = 1 - \prod_{i=1}^{n}[1 - P_i]$$

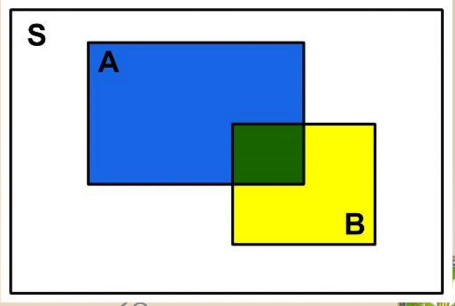
One minus the probability of no PFMs occurring

Uni-Modal Bounds

- For *n* positively dependent events, the probability of the union can be bounded as follows:
  - Lower bound, perfectly correlated
  - Upper bound, statistically independent events
  - If one event is dominant, the upper and lower bounds are about equal, weakest link controls

$$\max_{i} P(E_i) \le P(E) \le 1 - \prod_{i=1}^{n}[1 - P(E_i)]$$





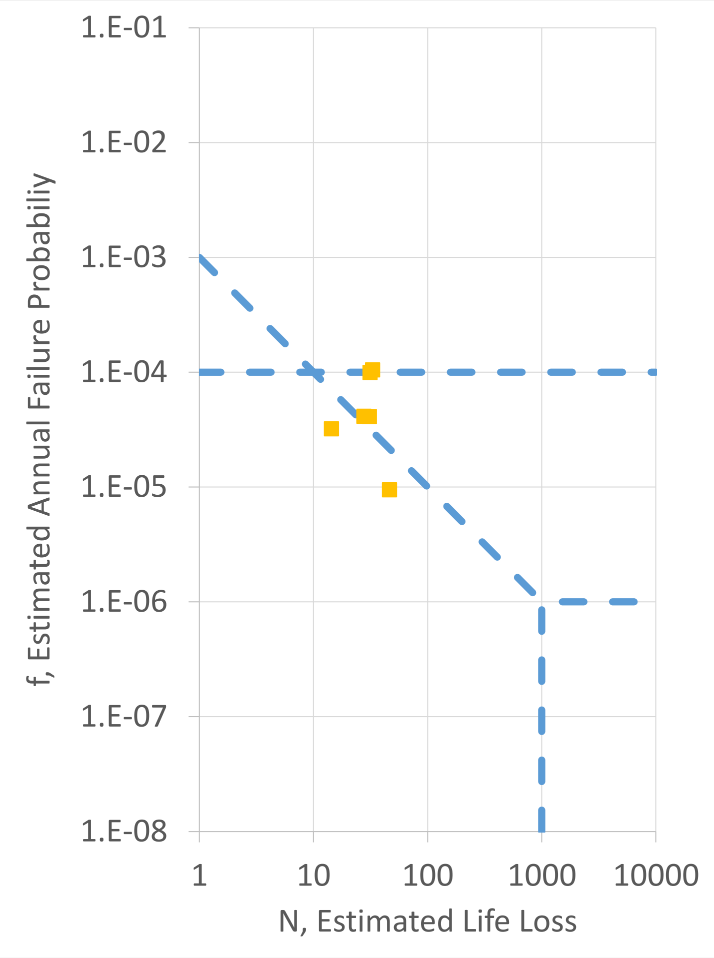
Monte Carlo Simulation

- Used to evaluate output uncertainty
- When analytical solutions are difficult (or don't exist)
- An output distribution is built up over thousands of simulation trials
- Basic Steps:
    - Build a model or event tree
    - Assign probability distributions to the model inputs
    - Sample the model inputs based on their probability distributions
    - Record the output(s)
    - Evaluate the probability distributions of the model output(s)

Monte Carlo Example

- AFP = P(Flood) * P(Failure | Flood)
- ALL = AFP * Life Loss

P(Flood)	P(Failure|...)	Life Loss	AFP	ALL
4.1E-4	0.10	30.9	4.1E-5	1.3E-3
5.1E-4	0.081	27.7	4.1E-5	1.1E-3
9.5E-4	0.11	33.0	1.0E-4	3.4E-3
1.4E-3	0.071	31.3	9.9E-5	3.1E-3
2.7E-4	0.035	46.5	9.5E-6	4.4E-4
6.3E-4	0.051	14.4	3.2E-5	4.6E-4



Conclusion

- Risk analysis is based on fairly simple set theory, probability theory, and statistical concepts.
- Risk analysis should not be viewed as a “black box”. Understanding what is happening mathematically is well within the ability of most risk analysis participants.
- The formulas used are “exact”. However, the outputs are only as good as the inputs (which are uncertain), so the outputs should not be interpreted as exact numbers.
- Ensuring that the right conceptual model is being used is more important than striving for numerical precision.

Combining Probabilities Example

For a gravity dam potential failure mode comprised of the following four events, estimate the annualized failure probability

1. Event F – A flood overtops the dam, $p(F) = 0.00002$
2. Event I – Foundation erosion initiates at the toe of the dam, $p(I|F) = 0.6$
3. Event E – Foundation erosion progresses and causes a weak plane to daylight, $p(E|F \cap I) = 0.2$
4. Event B – Dam breaches due to sliding instability $p(B|F \cap I \cap E) = 0.8$